

Galaxy Zoo Supernovae[★]

A. M. Smith,^{1†} S. Lynn,¹ M. Sullivan,^{1†} C. J. Lintott,¹ P. E. Nugent,² J. Botyanszki,²
M. Kasliwal,³ R. Quimby,³ S. P. Bamford,⁴ L. F. Fortson,⁵ K. Schawinski,^{6,7†}
I. Hook,^{1,8} S. Blake,¹ P. Podsiadlowski,¹ J. Jönsson,¹ A. Gal-Yam,⁹ I. Arcavi,⁹
D. A. Howell,^{10,11} J. S. Bloom,¹² J. Jacobsen,² S. R. Kulkarni,³ N. M. Law,¹³
E. O. Ofek^{3†} and R. Walters¹⁴

¹*Department of Physics (Astrophysics), University of Oxford, DWB, Keble Road, Oxford OX1 3RH*

²*Computational Cosmology Center, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*

³*Cahill Center for Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA*

⁴*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD*

⁵*School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA*

⁶*Department of Physics, Yale University, New Haven, CT 06511, USA*

⁷*Yale Center for Astronomy and Astrophysics, Yale University, PO Box 208121, New Haven, CT 06520, USA*

⁸*INAF-Osservatorio di Roma, via Frascati 33, I-00040 Monteporzio Catone (Roma), Italy*

⁹*Department of Particle Physics and Astrophysics, Faculty of Physics, The Weizmann Institute of Science, Rehovot 76100, Israel*

¹⁰*Las Cumbres Observatory Global Telescope Network, 6740 Cortona Dr, Suite 102, Goleta, CA 93117, USA*

¹¹*University of California, Santa Barbara, Broida Hall, Mail Code 9530, Santa Barbara, CA 93106-9530, USA*

¹²*Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA*

¹³*Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St George Street, Toronto M5S 3H4, Ontario, Canada*

¹⁴*Caltech Optical Observatories, California Institute of Technology, Pasadena, CA 91125, USA*

Accepted 2010 November 5. Received 2010 November 5; in original form 2010 September 29

ABSTRACT

This paper presents the first results from a new citizen science project: Galaxy Zoo Supernovae. This proof-of-concept project uses members of the public to identify supernova candidates from the latest generation of wide-field imaging transient surveys. We describe the Galaxy Zoo Supernovae operations and scoring model, and demonstrate the effectiveness of this novel method using imaging data and transients from the Palomar Transient Factory (PTF). We examine the results collected over the period 2010 April–July, during which nearly 14 000 supernova candidates from the PTF were classified by more than 2500 individuals within a few hours of data collection. We compare the transients selected by the citizen scientists to those identified by experienced PTF scanners and find the agreement to be remarkable – Galaxy Zoo Supernovae performs comparably to the PTF scanners and identified as transients 93 per cent of the ~ 130 spectroscopically confirmed supernovae (SNe) that the PTF located during the trial period (with no false positive identifications). Further analysis shows that only a small fraction of the lowest signal-to-noise ratio detections ($r > 19.5$) are given low scores: Galaxy Zoo Supernovae correctly identifies all SNe with $\geq 8\sigma$ detections in the PTF imaging data. The Galaxy Zoo Supernovae project has direct applicability to future transient searches, such as the Large Synoptic Survey Telescope, by both rapidly identifying candidate transient events and via the training and improvement of existing machine classifier algorithms.

Key words: methods: data analysis – surveys – supernovae: general.

1 INTRODUCTION

Supernovae (SNe) have a profound influence upon many diverse areas of astrophysics. They are the key source of heavy elements in the Universe, driving cosmic chemical evolution. Their energy input can initiate episodes of star formation and they are themselves the product of the complex physics underlying the final stages of stellar evolution. The homogeneous nature of the thermonuclear

[★]This publication has been made possible by the participation of more than 10 000 volunteers in the Galaxy Zoo Supernovae project (<http://supernova.galaxyzoo.org/authors>).

†E-mail: arfon.smith@astro.ox.ac.uk (AMS); sullivan@astro.ox.ac.uk (MS)

‡Einstein Fellow.

Type Ia SNe provides the most mature and direct probe of dark energy. Despite this importance in astrophysics, we understand surprisingly little about the physics governing SN explosions. Only the progenitors of the core-collapse Type II SNe have been directly identified: the physical nature of other SN types remains uncertain (for reviews see Hillebrandt & Niemeyer 2000; Smartt 2009). We remain ignorant about many aspects of SN rates, light curves, spectra, demographics, and the dependence of these properties on the environment, progenitor composition and explosion physics.

In part, this is due to the historical difficulty and technical challenges associated with locating SNe in the required numbers to create statistically meaningful samples, particularly at low redshift, where high-quality follow-up data can most easily be attained. This situation has changed with the availability of large format CCD detectors. Automated, wide-field transient searches on dedicated 1–2 m class telescopes and facilities are underway, typically observing thousands of square degrees every few days (e.g. Keller et al. 2007; Law et al. 2009). These flux-limited ‘rolling searches’ select transient events without regard to host galaxy properties or type.

This large amount of imaging data naturally generates its own particular logistical challenges in dealing with the data flow and identifying transient astrophysical objects of interest in the data (‘candidates’) for scientific study and analysis. Of particular importance is the rapid identification of new candidates, once the imaging data have been obtained and processed. Though many aspects of survey operations, such as image processing, can be efficiently pipelined, the identification of new transient sources remains challenging, with human operators (‘scanners’) invariably charged with wading through new detections on a nightly basis. Though computer algorithms can assist with identifying objects of interest in the data, this scanning can still absorb a significant amount of researcher time. A related issue is spectroscopic follow-up, a limited resource that must be prioritized and allocated efficiently to the detected candidates, with the absolute minimum of false candidates observed.

Two high-redshift SN searches highlight these challenges. The Supernova Legacy Survey (SNLS; e.g. Astier et al. 2006) used the MegaCam instrument on the 3.6-m Canada–France–Hawaii Telescope to survey 4 deg² with a cadence of a few days. Following automated cuts on the signal-to-noise ratio and candidate shape, each square degree would typically generate ~200 candidates for each night of observation (Perrett et al. 2010). Visual inspection would decrease this number to ~20 plausible real transients. The Sloan Digital Sky Survey-II Supernova Survey (SDSS-SN; e.g. Frieman et al. 2008) used the SDSS 2.5-m telescope to survey a larger area of 300 deg², though to a shallower depth than the SNLS (Sako et al. 2008). After the removal of moving (solar system) objects, in the first season (3 month period), human scanners viewed 3000–5000 objects each night spread over six scanners (>100 000 over the whole season). Although this number was radically reduced in later seasons as more automated procedures were developed (~14 000 during season 2), the burden on human scanners was still large (Sako et al. 2008). With new wide-field transient surveys generating many more candidates than these two surveys, advances in both automated techniques and human scanning are clearly required.

This paper details a new method for sorting through SN candidates, based upon the citizen science project ‘Galaxy Zoo Supernovae’ (Lintott et al. 2008, 2011). New candidate transient events are uploaded to the Galaxy Zoo Supernovae website, and are vi-

usually examined and classified by members of the public, guided by a tutorial and associated decision tree. Each candidate is examined and classified by multiple people and given an average score, with the candidates ranked and made available for further investigation in real time. The advantages of this approach are considerable. First, the burden of candidate scanning is largely removed from the science team running the survey. Secondly, each candidate is inspected multiple number of times (versus once by a scanner in previous transient surveys), reducing the chances that the candidate could be missed. Thirdly, with a large number of people scanning candidates, more candidates can be examined in a shorter amount of time and with the global Zooniverse (the parent project of Galaxy Zoo Supernovae) user base this can be done around the clock, regardless of the local time zone the science team happens to be based in. This speed can even allow interesting candidates to be followed up on the same night as that of the SN discovery, of particular interest to quickly evolving SNe or transient sources. Fourthly, the large number of human classifications collected can be used to improve machine learning algorithms for automated SN classification.

This paper reports the results from the early operations (over ~3 month period) of this system. In Section 2, we describe the Palomar Transient Factory (PTF), data from which were used in the tests and running of Galaxy Zoo Supernovae. Section 3 describes Galaxy Zoo Supernovae, including the ranking system for candidates used by the citizen science classifiers. Section 4 has details of the tests and first results of the Galaxy Zoo Supernovae operation. We discuss the future direction of this project in Section 5.

2 THE PALOMAR TRANSIENT FACTORY

The PTF is a wide-field survey exploring the optical transient sky. The survey is built around the 48-inch Samuel Oschin telescope at the Palomar Observatory, recently equipped with the CFH12k mosaic camera (formerly at the Canada–France–Hawaii Telescope) offering a 7.8-deg² field of view and robotized to allow remote and automated observations. Observations are mainly conducted using the Mould-*R* filter.

A full description of the operations of the PTF experiment can be found in Law et al. (2009). Of most relevance for SN studies are the ‘5-d cadence’ and ‘dynamical cadence’ experiments, each using ~40 per cent of the observing time. The dynamic cadence revisits survey fields on time-scales of 1 min up to 5 d and is particularly sensitive to rapid transient events (as well as longer duration SNe), whereas the 5-d cadence is specifically targeted to extragalactic SN studies (Rau et al. 2009). Even in the 5-d cadence, images are typically taken in pairs separated in time by about 1 h. This is to help identify moving objects (i.e. asteroids) in the imaging data, which might otherwise masquerade as new transients.

2.1 PTF real-time operations

The PTF (near)-real-time search pipeline is hosted by the National Energy Scientific Computing Center (NERSC) at the Lawrence Berkeley National Laboratory (LBNL). After data are taken and transferred from the Palomar observatory to the NERSC, the pipeline generates new subtraction images within an hour (Nugent et al. 2010), subtracting an older, deep ‘reference’ image from the new observations. The two images are photometrically matched

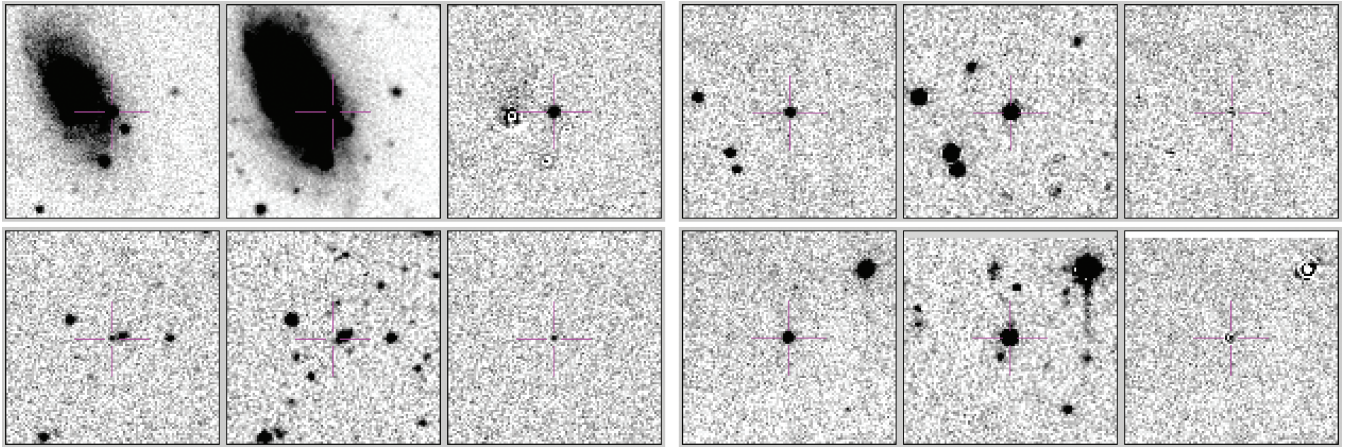


Figure 1. Four example detection triplets from the PTF, similar to those uploaded to Galaxy Zoo Supernovae. Each image is 100 arcsec on a side. In each triplet, the panels show (from the left-hand to right-hand side) the most recent image containing the candidate SN light (the science image), the reference image generated from data from an earlier epoch with no SN candidate light and the subtraction or difference image – the science image minus the reference image – with the SN candidate at the centre of the cross-hairs. The two triplets on the left-hand side are real SNe and were highly scored by the Zoo; the triplets on the right-hand side are not SNe and were given the lowest possible score.

using the *HOTPANTS* program,¹ an implementation of the Alard (2000) algorithm. Candidate transient events are then identified as $\geq 5\sigma$ detections in the subtraction images using *SEXTRACTOR* (Bertin & Arnouts 1996). Fluxes and various other relevant parameters are measured before storing all candidates in a data base. Each candidate is also ‘scored’ (producing the PTF ‘real-bogus’ value) using a machine-learning algorithm (the ‘PTF robot’) based on the characteristics of the detection and previous history of the candidate (Bloom et al., in preparation). The vast majority (~ 99.99 per cent) of these candidates are not real astrophysical transient events – the search algorithm is designed to be as inclusive as possible, with most of the candidates rejected via simple cuts. These include the following:

- (i) The ratio of both the semimajor and semiminor axes of the candidate shape to the seeing must be greater than 0.15 and less than 0.85, and the ratio of the full width at half-maximum (FWHM) of the candidate to the seeing must be greater than 0.5 and less than 2.0. These ensure that the candidate has a reasonable spatial extent when compared to the seeing.
- (ii) In a 7×7 pixel box placed on top of the candidate, the number of pixels deviating by more than 2σ must be less than six and the number deviating by more than 3σ must be less than two.
- (iii) Each candidate must be seen in at least one image taken in the previous 10 nights (including the night of detection), a constraint designed to remove fast moving solar system objects.
- (iv) Candidates within 1 arcsec of previously located objects (excluding the previous 10 nights) are removed to avoid the repeated detection of (e.g.) active galactic nuclei or variable stars.

The effectiveness of these cuts means that a typical full night of PTF observing will yield ~ 100 – 500 (average ~ 200) candidates that survive these culls, which can then be further sorted using only a short decision tree in Galaxy Zoo Supernovae.

Though the ultimate aim is to make the human scanners redundant with a fully automated machine-learning classification pipeline, at the present time, a substantial amount of human scanning is still

required to identify the good candidates (in part, this scanning can be used to train machine-based methods). Candidates are inspected visually by human scanners in the PTF team, using a web interface to reject false transient detections. The human scanner can dynamically alter a set of cuts to control the candidates that are shown for a given image, including the signal-to-noise ratio, shape parameters, the FWHM of the candidate compared to the global image value and the output score from the machine classifier. Based on the cuts chosen, the scanner is presented with a series of detection ‘triplets’ – each triplet contains three images showing the current image of the field (containing SN light together with all other objects), the historical or reference image of the same field (with no SN light) and the difference between the two (which should contain only the SN light). Examples of triplets are shown in Fig. 1. The human scanner then decides, based on his or her subjective (but informed) judgement, whether each of the candidates presented is a real transient event and, if so, then marks that candidate as either an SN-like transient or a variable star.

The primary goal of Galaxy Zoo Supernovae in the PTF is to initially supplement, but perhaps ultimately replace, the role of the PTF human scanners. By presenting a transient candidate to a number of different classifiers not only is the time of the PTF team freed to spend on tasks not suitable for the general public, but the potential of mis-classification of candidates due to individual human error is also significantly reduced. The 5-day and dynamical cadence programmes in the PTF collect data on every night of the year March to November (weather permitting) and on each night 2–4 of the PTF team share the scanning tasks, examining ~ 500 candidates. This not only requires several person-hours of work, but a large number of classifications by a small number of PTF scanners are also likely to contain errors and this is where the repeat classification by Galaxy Zoo Supernovae volunteers can help.

The Galaxy Zoo Supernova project also has other aims. A longer term goal is to provide sufficient classification data for the training and improvement of the PTF machine-learning classification algorithm. A final consideration is to build expertise in the citizen science community for future transient surveys, which of course generate many more candidates than the PTF, perhaps approaching thousands of genuine candidates on a nightly basis.

¹ <http://www.astro.washington.edu/users/becke/hotpants.html>

3 GALAXY ZOO SUPERNOVAE

3.1 Description of a typical ‘Zoo’

The Galaxy Zoo Supernovae website² is built using the ZOONIVERSE³ API (Application Programming Interface) toolset. The ZOONIVERSE API is the core software supporting the activities of all Zooniverse citizen science projects. Built originally for Galaxy Zoo 2, the software is currently being used by six different projects. The ZOONIVERSE API is designed primarily as a tool for serving a large collection of ‘assets’ (e.g. images or videos) to an interface and collecting back-user-generated interactions with these assets.

So that the project website can retain a high performance during spikes of activity, Galaxy Zoo Supernovae is hosted on the Amazon Web Services, which provides a virtualized machine environment that can autoscale in size based upon the server load. The site uses the Elastic Compute Cloud⁴ (EC2) for web/data base servers and the Simple Storage Service⁵ (S3) for image storage.

Image assets are presented to volunteers of the website through custom user interfaces, designed to aid the volunteer in classifying the object. For many projects, this interface takes the form of a decision tree, which walks the volunteer through a number of questions concerning the current image. The interaction of the volunteer with the website produces a set of ‘annotations’, which together constitute a ‘classification’ of the asset. These are stored for later analysis or in the case of Galaxy Zoo Supernovae are scored in real time to change the behaviour of the website.

3.2 Galaxy Zoo Supernovae website operations

Similar in nature to the original Galaxy Zoo 2 interface, Galaxy Zoo Supernovae is a classic example of a ‘Zoo’. When a new highly scored candidate is located in the PTF pipeline, an image triplet (Fig. 1) of the candidate is automatically uploaded, together with a small amount of metadata, to the Galaxy Zoo Supernovae API. Upon uploading, the image is saved to Amazon S3 (a file hosting service) and registered with the website. Finding new SNe is time-critical and our method of automatically registering new assets with the API means that classifiers are inspecting SN candidates discovered just hours earlier. The interface for Galaxy Zoo Supernovae presents these candidate detection triplets (just as with the PTF human scanners, Section 2.1) together with a decision tree of questions and answers designed to help classify each candidate (see Fig. 3). Fig. 2 displays the typical flow in the system. Once a candidate has been classified (see below), it is instantly available to the PTF team through a private web interface.

3.3 Decision tree

The decision tree developed to assist volunteers in classifying candidates is described in Fig. 3. This decision tree is designed to remove as many false candidates as possible, without losing real, scientifically interesting events. In this respect, the decision tree is conservative in the candidates that are removed to minimize the number of false negatives. The tree proceeds as follows:

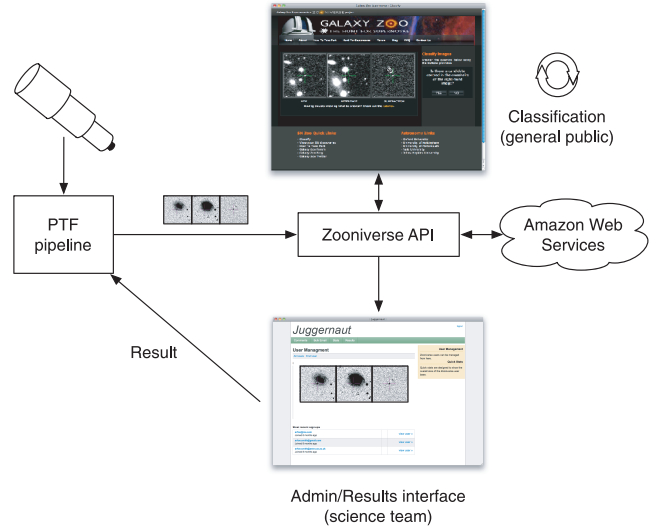


Figure 2. A schematic showing the data acquisition and analysis in Galaxy Zoo Supernovae: raw data are processed by the PTF pipeline, automatically uploaded to the API, presented, analysed and scored by the Zooniverse community, and available for review by the PTF science team.

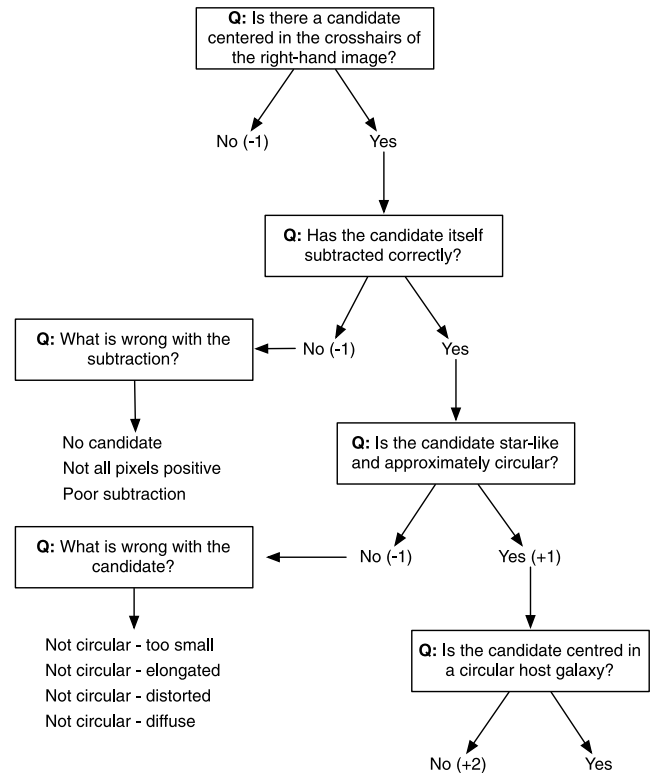


Figure 3. The decision tree that a Galaxy Zoo Supernovae volunteer is presented with when classifying a candidate (see Section 3.3). The decision tree can end at a number of points. The scoring points in the decision tree are also shown. Both the path through the decision tree and the cumulative score are recorded for later analysis.

(i) *Is there a candidate centred in the cross-hairs of the right-hand image?*

The PTF subtraction pipeline can occasionally undergo a failure and report (and therefore upload to the site) a ‘good’ candidate that is actually an error in the processing. This can be due to large

² <http://supernova.galaxyzoo.org/>

³ <http://zooniverse.org>

⁴ <http://aws.amazon.com/ec2>

⁵ <http://aws.amazon.com/s3>

(several pixel) misalignments of the two images being analysed, often localized in a particular part of the CCD where the astrometric solution fails. Other sources of failure include saturated pixels or bleed trails from bright stars, or problems with the pipeline flat-fielding. The *SETRACTOR* detection algorithm can also sometimes detect a noise peak rather than a real transient. Though the basic cuts made by the PTF remove most of these errors, on occasion they are ranked highly and uploaded to Galaxy Zoo Supernovae (emphasizing the need for human classifiers). Therefore, the first question in the decision tree is designed to remove such objects. The right-hand image in the triplets in Fig. 1 is the focus of this question.

(ii) *Has the candidate itself subtracted correctly?*

Small misalignments between the reference and science image can result in image-subtraction problems, usually indicated by a dipole of positive and negative pixels in the subtraction image. The cores of bright (but not saturated) stars can also mis-subtract and result in ‘bulls-eye’ patterns in the subtraction images. This question is designed to flag such candidates.

(iii) *Is the candidate star-like and approximately circular?*

This question is designed to remove unidentified cosmic rays or diffuse/non-circular candidates, which result from image-subtraction problems. The volunteer is asked if the candidate looks like a round, symmetrical dot (star). Candidates that are very small (1–2 pixels i.e. not point spread function like), elongated or otherwise distorted, or diffuse would trigger a negative response to this question.

(iv) *Is the candidate centred in a circular host galaxy?*

The final question is more subjective and is designed to categorize real astrophysical transients into two broad categories. Many of the transients, which the PTF detects, are variable stars lying within our own galaxy, which are of interest to a different set of science users than extragalactic transients. Variable star transients will appear to lie in ‘hosts’ that are circular (as they are stars) and will also appear to be located in the centre of these hosts. By contrast, SNe will either have no host galaxy or will lie (probably off-centre) in a large diffuse host galaxy. This question therefore broadly splits the real transients into variable stellar transients and SNe. Most SNe that do happen to lie in the centres of their host galaxies will not be categorized as variable stars – the question also requires the ‘host galaxy’ to be circular.

A full tutorial is available to new volunteers of the website to illustrate the different questions using real PTF data.

3.4 Asset scoring and priority

Once a volunteer has examined a candidate, their response is converted into a score, S , as follows:

- (i) The initial score is zero.
- (ii) If a classifier answers negatively, any question up to and including ‘Is the candidate star like and approximately circular?’, the candidate is given a score of -1 .
- (iii) If a classifier instead answers positively up to that question, then the candidate is given a score of $+1$.
- (iv) If the classifier then also marks the candidate as not centred in a circular host, then the candidate gains an *additional* score of 2.

The structure of the decision and scoring of the questions means that candidates can only end up with a score of -1 , 1 or 3 from each classification, with the most promising SN candidate’s score 3. As each new classification is received, the arithmetic mean score

(S_{ave}) of the candidate is recalculated. Candidates which are not astrophysically interesting tend to have $S_{\text{ave}} < 0$ (i.e. most volunteers scored them a ‘ -1 ’). Astrophysical transients typically have $S_{\text{ave}} > 0$ and SNe tend to have $S_{\text{ave}} > 1$ (i.e. most volunteers scored them a ‘ 3 ’).

The asset prioritization system is adjusted after each classification is received and operates to prioritize the best SN candidates (i.e. the order in which the candidates are shown to classifiers). When new candidates are uploaded to the website, they are initially prioritized based upon (i) a score supplied by the PTF pipeline; and (ii) the age of the candidate (the newest uploads are shown first). The PTF ‘real bogus’ value (Section 2.1) is calculated by the PTF pipeline for all candidates and gives an indication of the likelihood that a candidate is a real transient. This value is only used to determine the order in which candidates are shown and is not used in the final ranking.

Studies of results from early (‘beta’) versions of Galaxy Zoo Supernovae have allowed us to optimize the asset prioritization to reduce the time taken to identify candidates. We divide candidates into the following four categories:

- (i) *Unseen* – candidates which have three or fewer classifications.
- (ii) *Bulk* – candidates which have been classified between three and 10 times.
- (iii) *Stragglers* – candidates which have been classified more than 10 times, but which do not have a ‘definitive’ S_{ave} (i.e. those with $0.0 < S_{\text{ave}} < 1.7$).
- (iv) *Done* – candidates which have been classified more than 10 times and which have $S_{\text{ave}} < 0.0$ or $S_{\text{ave}} > 1.7$, and candidates which have been classified more than 20 times.

Candidates in the ‘unseen’ category are given absolute precedence over all others in an aim to get an initial understanding of the quality of the candidate; they are shown in order of upload time followed by the real-bogus score. Once these are completed, the ‘bulk’ and ‘straggler’ candidate classes have equal priority. We select randomly between the two classes, choosing the newest candidate with the highest score from each group – as a candidate begins to receive ‘positive’ classifications (i.e. S of 1 or 2), it is prioritized above any others, thus allowing rapid identification of the most interesting targets.

The choice of 10 classifications as the first point at which a candidate can be considered classified is a compromise between the robustness of the classification and speed. Clearly, the greater the number of classifications required for each candidate, the slower the classification process proceeds; yet the process must be robust against both user mistakes (i.e. clicking the wrong button) and misunderstanding.

The aim is both to quickly classify the best, high-scoring candidates (which will rapidly exceed the $S_{\text{ave}} = 1.7$ threshold after 10 classifications) and to remove the worst candidates (which will remain below $S_{\text{ave}} = 0.0$). More ambiguous candidates can then obtain up to 10 extra classifications before completion. The process continues until a target has received enough classification scores that it is considered ‘done’, at which point it is removed from the pool of available candidates. Our simulations based on the beta versions indicated that this scheme is two to three times faster at classifying than just using a random order.

3.5 Communication of results

The science of Galaxy Zoo Supernovae relies on new candidates being classified rapidly and those classifications then being easily accessible to the science team.

3.5.1 Science dashboard

A key part of the Galaxy Zoo Supernovae website is a science ‘dashboard’ for the PTF team. The science dashboard provides basic statistics on the number of candidate uploads, classifications and volunteers versus time as well as a more in-depth breakdown of the classification history for a candidate or individual.

Custom views have been created, which break down a score-ranked list of candidates for each day and week, allowing observing teams to use these rankings to help in the identification of good candidates for follow-up observations. Candidates already identified as PTF transients show the PTF identifier on the science dashboard and a link is also provided to allow the science team to easily mark a highly ranked candidate from the Zoo in the PTF data base.

3.5.2 Candidate alerts

In order to improve the rate at which objects are classified, an automated alert system that monitors the number of candidates being uploaded to the website is used. Should the number of unclassified candidates reach a threshold, the website sends an automated ‘alert’ to Galaxy Zoo Supernovae subscribers. These (email) alerts are usually sent out once per day, coinciding with the end of a night’s candidates being uploaded from the NERSC, and usually result in the full complement of candidates being classified within a few hours.

3.5.3 ‘My Supernovae’

Providing feedback to the Galaxy Zoo Supernovae community is a vital part of the overall website experience to encourage volunteers to return to the website. This is partly done using forums and blogs, where scientists can comment on individual events classified by the Zoo. In addition, each volunteer can view a history of the candidates that they have classified on their ‘My Supernovae’ (MySN) page.

The MySN page displays the candidate triplets. Those which have been observed are overlaid with a small symbol identifying the candidate as an SN, variable star or asteroid. Clicking on one of the candidates also allows the volunteer to see the average rating across all classifications, the number of classifiers and whether the candidate was selected for follow-up by the PTF team. PTF observers are encouraged to leave comments on the science dashboard that the classifiers can also see on their MySN page.

4 RESULTS

Galaxy Zoo Supernovae was first trialled on two specific occasions supporting PTF spectroscopic follow-up observations at the 4.2-m William Herschel Telescope (WHT), in 2009 August and October. The selection of the candidates observed by the WHT was guided by the Zoo results, with a particular emphasis on comparing the classifications produced by Galaxy Zoo Supernovae with those produced by PTF human scanners working on the same data. The top 20 scored candidates from this initial trial run of Galaxy Zoo Supernovae are shown in Fig. 4. 16 of these candidates were observed by the WHT; 15 were confirmed as SNe, with one cataclysmic variable.

Since 2010 April, Galaxy Zoo Supernovae has been running full-time on PTF candidates and by 2010 July 15 had classified $\approx 13\,900$ SN candidates at the rate of several hundred candidates per observing night. In all but the earliest weeks of the project, all submitted candidates were classified by the Zoo. This classified

sample forms the basis of our analysis in this section. A distribution of the scores (S_{ave}) for all of these candidates can be found in Fig. 5. The bulk of the candidates uploaded are classified as likely not astrophysically real events and correspond to subtraction artefacts or other reduction problems. This is indicative of the conservative cuts that are made in the PTF pipeline to avoid losing real SN events for follow-up and highlights the currently essential requirement for visual inspection of the pipeline candidates.

4.1 Comparison with professional classifiers

The performance of the public at classifying candidates can be gauged by comparing with the classifications the PTF team assigned to the same objects. The PTF team broadly classify objects into four visual categories: not interesting (not assigned a type), asteroids, variable stars and transients (such as SNe). Asteroids are not screened for by Galaxy Zoo Supernovae – only one image is uploaded for each PTF candidate, which clearly cannot be used to distinguish moving objects. Asteroids are typically removed from the candidate list prior to upload by insisting on two separate detections of a candidate within 1 arcsec of each other, though this process is not perfect, particularly with slow-moving asteroids where the apparent motion can only be a few arcseconds a day.

To illustrate the performance of Galaxy Zoo Supernovae, we split the candidates by their PTF-assigned categories and calculate the fraction in each category as a function of S_{ave} . Fig. 6 is a stacked box plot of the results. At low scores, practically, all candidates are those which the PTF team decide are not interesting: these will include poor subtractions, artefacts/cosmic rays, etc. As S_{ave} increases we see a steady rise in the number of both variable star candidates and transients. By a score of around 1.4, variable stars are no longer selected and instead the majority of the candidates are SN-like transients.

A number of caveats should be borne in mind when examining this plot. The first is that not all variable stars identified by Galaxy Zoo Supernovae will be assigned that type by the PTF scanners. As the primary goal of the PTF is the study of explosive transients, variable stars are frequently not recorded in the PTF catalogue (i.e. they will be assigned ‘No type’ in Fig. 6). The second caveat is that each PTF candidate is potentially observed many times over a period of several weeks over many epochs, yet should only be uploaded to Galaxy Zoo Supernovae once. If there is some problem with the particular epoch that is uploaded to the Zoo (a poor image subtraction or poor seeing conditions), then a real astrophysical event may be poorly scored by the Zoo on that epoch. However, that candidate may potentially be saved by a human scanner based on an image from a different epoch. Thus, real transient events can occasionally be poorly scored by the Zoo if the uploaded image is of poor quality; this is the case for some of the real transients that scored $S_{\text{ave}} < 0$. Finally, it is important to note that the true nature of many of the candidates remains unknown, and the comparison drawn here is between the Zoo selection and that of a subjective (though experienced) expert opinion.

Fig. 6 demonstrates that Galaxy Zoo Supernovae is capable of prioritizing good candidates and that the highest ranked candidates are likely to be SNe rather than variable stars. The candidates which were classified as asteroids in the Galaxy Zoo Supernovae sample are given a relatively high score by the Zoo volunteers – they typically mimic high-quality ‘host-less’ transient events.

Some of the Galaxy Zoo Supernovae classified candidates were observed spectroscopically by the PTF collaboration, as well as candidates identified by other techniques. We examine the S_{ave}

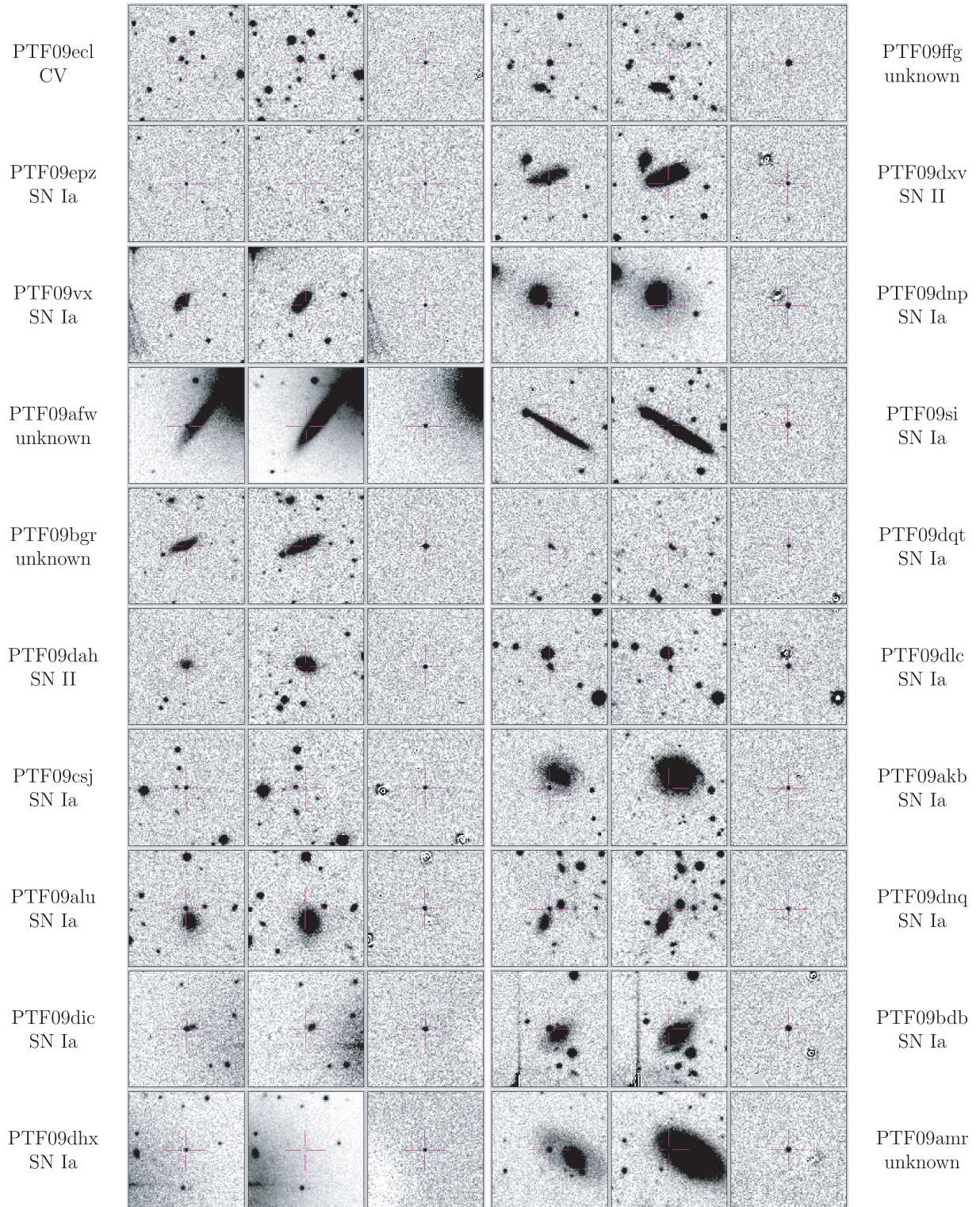


Figure 4. A montage of the 20 highest ranked PTF candidates from the October testing of the website. Each set of three images shows, from the left-hand to right-hand side, the new image, the reference image and the subtraction image. The position of the candidate is shown in each panel by the cross-hairs. The candidate name and the spectroscopic type from the WHT (where available) are also shown.

distribution for these ~ 140 spectroscopically confirmed SNe (Fig. 7), equivalent to approximately five to six full nights of 4-m-class telescope time (spread over 10 actual nights with a mix of screening and follow-up of previously confirmed transients). Approximately, 93 per cent of these SNe gathered by the PTF over

2010 April–July were highly scored ($S_{\text{ave}} > 0$) by the Zoo (60 per cent have $S_{\text{ave}} > 1$) and real SNe with an $S_{\text{ave}} < 0$ comprise only 0.1 per cent of all Zoo objects scored with $S_{\text{ave}} < 0$. Though this may represent a slightly biased test (low-scored candidates are less likely to be followed spectroscopically), there are other techniques

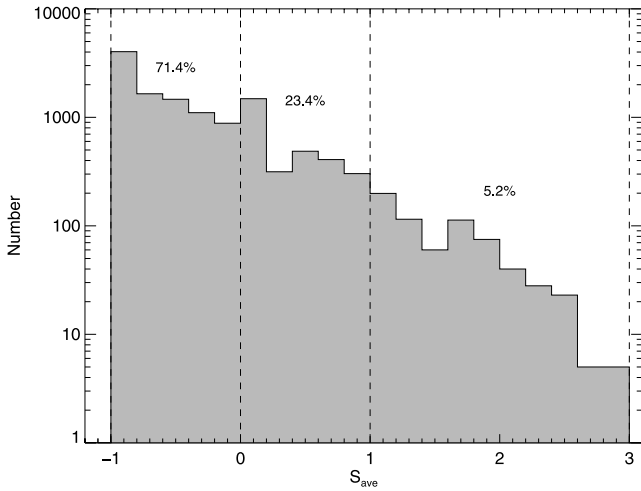


Figure 5. The distribution of all of the scores (S_{ave}) for all of the PTF candidates classified by Galaxy Zoo Supernovae between 2010 April–July. $\approx 13\,900$ candidates were classified. The bulk of these ~ 70 per cent – were classified as not astrophysically real by the Zoo ($S_{\text{ave}} < 0$). Only one in 20 candidates was identified as likely SN event.

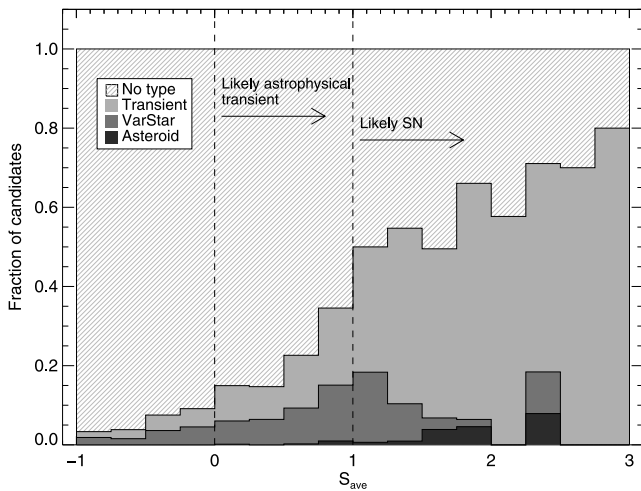


Figure 6. A breakdown of the classifications collected during operations of Galaxy Zoo Supernovae. The bars show the distribution of candidate types (as determined by the PTF team) for a given Zoo score (S_{ave}). The PTF team potentially assign a classification of asteroids, variable stars or transients to each Zoo candidate – objects without a PTF classification are deemed not to be interesting. Galaxy Zoo Supernovae is not designed to flag moving objects, which are largely removed before upload. Note that not all variable stars will be saved to the PTF data base, so this category is likely highly incomplete (and the variables stars will appear as ‘No type’).

for screening candidates within the PTF that complement the Zoo that partially mitigate this bias. These include the PTF robot (Section 2.1) and some human scanning effort. It is encouraging that, to the degree that we can test it, the Zoo is capable of selecting the best SNe from the PTF data set. We also note that no highly scored Zoo candidate that was observed spectroscopically turned out not to be an SN.

Note that the majority of the highly scored Galaxy Zoo Supernovae remain unobserved spectroscopically, particularly at the fainter end of the candidate brightness. Some of these candidates are asteroids (or more accurately, objects that only appear on a single

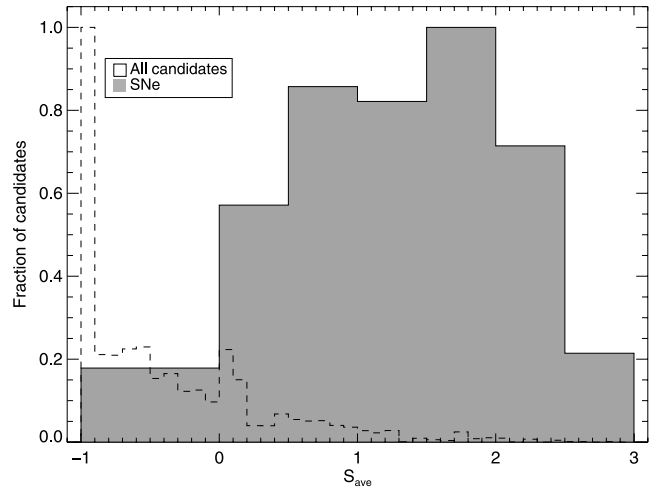


Figure 7. A breakdown of the scores (S_{ave}) for the 140 known SNe identified via PTF follow-up spectroscopy (grey histogram). For reference, the distribution of the S_{ave} measures for all the objects is shown as the open histogram. These classifications were collected during 2010 April–July.

night of data). Others are probably real SNe for which there was insufficient follow-up time available.

4.2 Effect of candidate brightness

Fig. 8 shows candidate scores from Galaxy Zoo Supernovae as a function of the photometric apparent R magnitude of the candidate (with the host light subtracted) and the magnitude error, both taken from the P48 PTF search pipeline. We plot these relations separately for spectroscopically confirmed SNe and PTF transients and show the comparison with all PTF candidates as a set of contours. The latter comparison highlights just what a small fraction of all the PTF candidates, the real SNe and transients represent.

Fig. 8 shows a few interesting trends. For the confirmed SNe, there is a mild decrease in S_{ave} as the candidates become fainter (or have a larger error), at about $\sim 3\sigma$ significance or $\sim 6\sigma$ when considering the magnitude error. (There is an equivalent trend for all the PTF transients.) This is expected – at fainter magnitudes, SNe become harder to identify visually with a noisier detection and the classification becomes more subjective. The SNe are also likely to be at higher redshift, and thus perhaps appear more centrally located in fainter host galaxies and are more likely to fail the final step in the decision tree (Fig. 3).

None the less, Galaxy Zoo Supernovae clearly identifies and scores highly the bulk of the SNe from the PTF, and at bright to intermediate magnitudes, the separation of SNe is robust. Even at fainter magnitudes, the majority of the SNe score $S_{\text{ave}} > 0$ and above a detection significance of $\sim 8\sigma$, the Zoo scores all SNe and the vast majority of PTF transients at $S_{\text{ave}} > 0$.

4.3 The scoring model

An analysis of the scoring model can reveal optimizations that can be made to the number of classifications required for each candidate. As an example, we show the ‘trajectory’ of S_{ave} for PTF candidates as a function of the number of classifications in Fig. 9. As expected, the variation in S_{ave} when adding additional classifications is larger when the total number of classifications is small compared to when many classifications are available. It is also apparent that once ~ 15

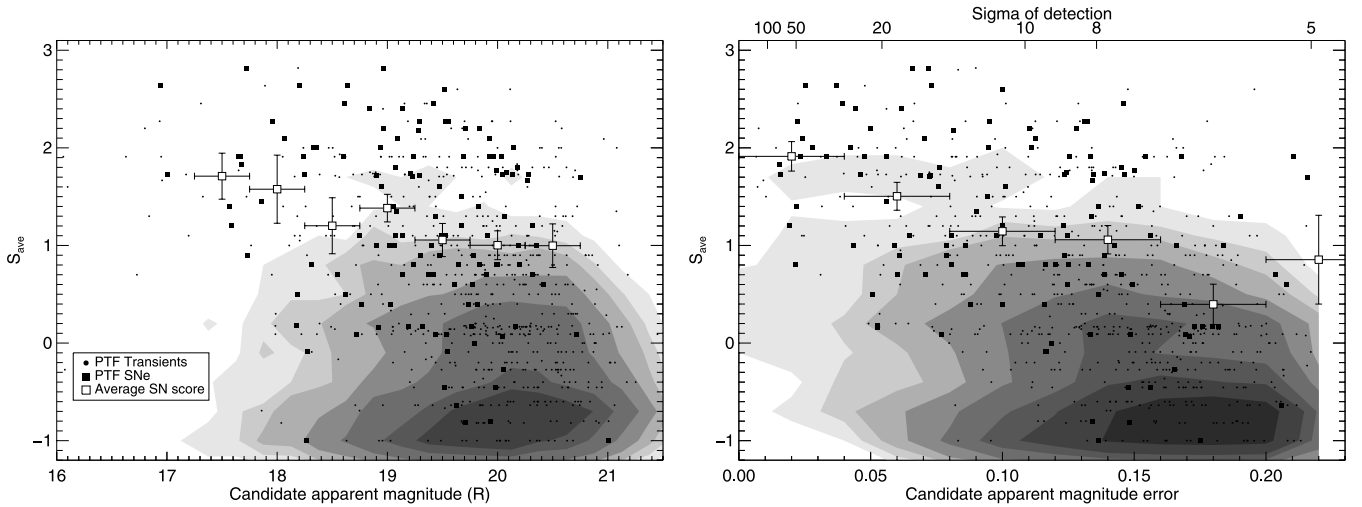


Figure 8. The Galaxy Zoo Supernova scores S_{ave} of PTF candidates of various types as a function of their apparent R detection magnitude (left-hand panel) and the error in that magnitude (right-hand panel). The filled circles show PTF objects believed to be SN-like transients and filled squares show the confirmed SNe, while the contours show the distribution of all $\sim 14\,000$ PTF candidates. The open squares show the average SN scores in bins of magnitude (or magnitude error). For these candidates, the trend of decreasing score with increasing magnitude is significant at about 3σ and with increasing magnitude error at $\sim 6\sigma$. Note that only detections of 5σ significance or greater are uploaded to the Zoo, hence the cut-off in the right-hand panel.

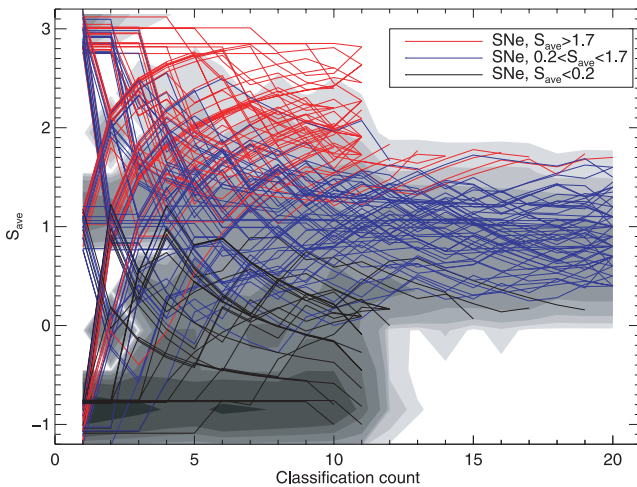


Figure 9. The Galaxy Zoo Supernova scores S_{ave} of PTF candidates of various types as a function of the number of classifications they have received. Each line represents a spectroscopically confirmed PTF SN. Those in red have a final $S_{\text{ave}} > 1.7$, those in black a final $S_{\text{ave}} < 0.2$ and those in blue intermediate scores. Only candidates scored with $0.2 < S_{\text{ave}} < 1.7$ continue to be classified beyond 10 classifications. The grey contours show the trajectories of all PTF candidates that were classified by the Zoo, regardless of any spectroscopic typing. As the scores are highly quantized (each classification can only result in a score of -1 , 1 or 3), each line representing a PTF SN is offset slightly in S_{ave} for clarity.

classifications have been received, very few candidates change S_{ave} significantly.

We also examine the dispersion in each of the Galaxy Zoo Supernova scores, as calculated from the individual classifications, as a function of the scores themselves. Fig. 10 plots the mean absolute deviation in the score of each classified candidate as a function of the final candidate score. As the individual scores from which each S_{ave} is calculated are highly quantized (each classification can only result in a score of -1 , 1 or 3), the resulting plot is highly structured. In particular, objects with S_{ave} of -1 or 3 must have a dispersion of zero and a further dip in the dispersion is also seen around the third

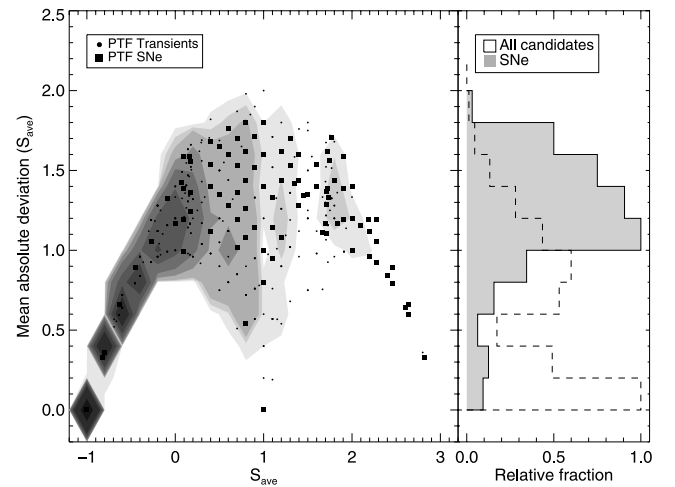


Figure 10. The dispersion (mean absolute deviation) of the Galaxy Zoo Supernova scores as a function of the scores themselves (S_{ave}) for PTF candidates of various types. The filled circles show PTF objects believed to be SN-like transients and filled squares show the confirmed SNe, while the contours show the distribution of all $\sim 14\,000$ PTF candidates. The histograms show the distribution of the mean absolute deviations in S_{ave} for confirmed SNe and for all PTF candidates. In principle, with a more refined scoring grid, the dispersion in each candidate score could be used as a measure of the confidence of the final classification, but our current simple decision tree does not permit this.

scoring possibility, 1. While, in principle, the dispersion in the score might be thought of as a good measure of the classification confidence (measuring, in essence, the agreement between individual classifiers), the current simple decision tree is not refined enough to allow this statistic to be useful.

Therefore, there exists some room to improve the scoring model used by Galaxy Zoo Supernovae (and hence the efficiency of the project). A detailed analysis of the data in Fig. 9 shows that reducing the number of classifications needed before a candidate is considered classified (Section 3.4) from >20 to >15 (for intermediate scoring events) and from >10 to >8 (for low and high scoring

events) would reduce the total number of classifications recorded by ~ 20 per cent, while only moving a handful of candidates (< 5 per cent) across the boundaries of $S_{\text{ave}} = 1.7$ and 0.2 .

In principle, an analysis of which volunteers consistently get the classifications correct (when compared to a professional astronomer or a spectroscopic classification) could be used to weight different volunteer responses. For example, an experienced classifier with a consistent history of correct responses could have a larger weight than a novice volunteer – there is evidence from Fig. 9 that even good SN candidates can receive the lowest score (many SN trajectories start at an $S_{\text{ave}} = -1$). Such a feature is not yet implemented in Galaxy Zoo Supernovae, but could be used to arrive at a final S_{ave} more quickly.

4.4 Volunteer behaviour

To date, over 13 000 individuals from the Zooniverse community have visited the Galaxy Zoo Supernovae website and 2800 have classified one or more SN candidates. This project relies upon the rapid classification of SN candidates and although the community is relatively small compared to, for example, Galaxy Zoo, a combination of email alerts and a committed core of a few hundred individuals has made Galaxy Zoo Supernovae a success.

An analysis of the fraction of classifications contributed compared to the average number of classifications per user shows that close to 90 per cent of the classifications in Galaxy Zoo Supernovae are contributed by less than 20 per cent of the community. In the Galaxy Zoo 2 project (Masters et al. 2010; Lintott et al., in preparation), close to 50 per cent of the classifications were by individuals whose total classification count was less than 10 galaxies; for Galaxy Zoo Supernovae that fraction is 3 per cent.

5 FUTURE DIRECTIONS

This paper has introduced Galaxy Zoo Supernovae, a new web-based citizen science project modelled after ‘Galaxy Zoo’, which uses members of the public to identify good SN candidates from wide-field imaging data. Using data from the PTF, we have shown that the citizen scientists are extremely good at identifying real SNe from amongst the thousands of candidates that the PTF generates, with only a small ‘false negative’ rate at the faintest candidate magnitudes.

Clearly, Galaxy Zoo Supernovae is not restricted to PTF data and can, in principle, be applied to any future imaging survey, such as SkyMapper (e.g. Keller et al. 2007) or Pan-STARRS-1 (e.g. Kaiser 2004). The candidate upload mechanism is flexible, and the triplet format (Fig. 1) simple, with custom result pages easily produced for individual surveys. Perhaps the most exciting aspect for massive future transient surveys, such as the Large Synoptic Survey Telescope (LSST), will be the use of Galaxy Zoo Supernovae classification data to improve the training and accuracy of automated machine-learning transient classifiers (Starr et al. 2009).

The underlying concept of Galaxy Zoo Supernovae is easily extended. For example, there is also no need to restrict the project to single images of new transient events. Multiple images of a potential SN from different epochs, that is, a candidate history, could also be uploaded to improve the accuracy of the classifications and thus reduce the possibility of a misclassification due to a single poor subtraction. If this included data from before the candidate was first detected, those candidates with a history of poor subtractions could quite trivially be eliminated. Those asteroids and moving objects which do get uploaded could also be removed by visually comparing

the candidate position on several epochs. Galaxy Zoo Supernovae could also be used to identify new transients triggered by detections at other wavelengths, for example, to quickly identify optical counterparts to gamma-ray bursts, where previous optical reference images might not exist and a timely search is critical for follow-up.

Galaxy Zoo Supernovae could also be used for precise volumetric SN (or any transient) rate determinations. In these calculations, the efficiency of the search (the ratio of recovered to actual SN events) needs to be accurately known, as a function of apparent magnitude and other SN properties. By uploading ‘fake’ candidates (artificial SN events inserted into the images) as well as real SNe, the reliability of the Zoo can be determined accurately and allow the discovery rate to be converted into a real physical SN rate.

With the discovery stream of new transient types becoming ever larger, and the dramatic increase set to continue with future surveys, such as the LSST, the burden of identifying the best new candidates increases correspondingly. By engaging the considerable interest and enthusiasm of the public, we have demonstrated that citizen science projects, like Galaxy Zoo Supernovae, can play a major role in ongoing and future transient surveys.

ACKNOWLEDGMENTS

We acknowledge the valuable contributions of the Zooniverse community without which this project would not have been possible. AMS acknowledges support from the Leverhulme Trust. MS acknowledges support from the Royal Society. MS and AG-Y acknowledge support from a Weizmann–UK ‘Making connections’ grant. CJL acknowledges support from the STFC Science in Society Program and The Leverhulme Trust. PEN acknowledges support from the US Department of Energy Scientific Discovery through Advanced Computing program under contract DE-FG02-06ER06-04. KS acknowledges support from a NASA Einstein Postdoctoral Fellowship grant number PF9-00069, issued by the Chandra X-ray Observatory Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of the NASA under contract NAS8-03060. JSB acknowledges support of an NSF-CDI grant ‘Real-time Classification of Massive Time-series Data Streams’ (Award #0941742). The National Energy Research Scientific Computing Center, which is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231, provided staff, computational resources and data storage for this project.

The William Herschel Telescope is operated on the island of La Palma by the Isaac Newton Group in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias. Observations obtained with the Samuel Oschin Telescope at the Palomar Observatory as part of the PTF project: a scientific collaboration between the California Institute of Technology, Columbia University, Las Cumbres Observatory, the Lawrence Berkeley National Laboratory, the National Energy Research Scientific Computing Center, the University of Oxford and the Weizmann Institute of Science.

REFERENCES

- Alard C., 2000, *A&AS*, 144, 363
- Astier P. et al., 2006, *A&A*, 447, 31
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Frieman J. A. et al., 2008, *AJ*, 135, 338
- Hillebrandt W., Niemeyer J. C., 2000, *ARA&A*, 38, 191

- Kaiser N., 2004, in Oschmann J. M., Jr, ed., SPIE Conf. Ser. Vol. 5489, Ground-based Telescopes. SPIE, Bellingham, p. 11
- Keller S. C. et al., 2007, PASA, 24, 1
- Law N. M. et al., 2009, PASP, 121, 1395
- Lintott C. J. et al., 2008, MNRAS, 389, 1179
- Lintott C. J. et al., 2011, MNRAS, 410, 166
- Masters K. L. et al., 2010, MNRAS, doi:10.1111/j.1365-2966.2010.17834.x
- Nugent P. et al., 2010, The Astronomer's Telegram, 2600, 1
- Perrett K. et al., 2010, AJ, 140, 518
- Rau A. et al., 2009, PASP, 121, 1334
- Sako M. et al., 2008, AJ, 135, 348
- Smartt S. J., 2009, ARA&A, 47, 63
- Starr D. L., Bloom J. S., Brewer J. M., Butler N. R., Poznanski D., Rischard M., Klein C., 2009, in Bohlender D. A., Durand D., Dowler P., eds, ASP Conf. Ser. Vol. 411, Astronomical Data Analysis Software and Systems. Astron. Soc. Pac., San Francisco, p. 493

This paper has been typeset from a $\mathrm{T}_{\mathrm{E}}\mathrm{X}/\mathrm{L}^{\mathrm{A}}\mathrm{T}_{\mathrm{E}}\mathrm{X}$ file prepared by the author.